CSAE WPS/2002-09

Do African Manufacturing Firms Learn from Exporting?

ARNE BIGSTEN[(1)], PAUL COLLIER[(2), (3)], STEFAN DERCON[(2)], MARCEL FAFCHAMPS[(2)], BERNARD GAUTHIER[(4)], JAN WILLEM GUNNING[(5)], ABENA ODURO[(6)], REMCO OOSTENDORP[(5)], CATHERINE PATTILLO[(7)], MÅNS SÖDERBOM[(2)], FRANCIS TEAL[(2)] and ALBERT ZEUFACK[(3)]

September 2002

Abstract

In this paper, we use firm-level panel data for the manufacturing sector in four African countries to estimate the effect of exporting on efficiency. Estimating simultaneously a production function and an export regression that control for unobserved firm effects, we find both significant efficiency gains from exporting, supporting the learning-by-exporting hypothesis, and evidence for self-selection of more efficient firms into exporting. The evidence of learning-by-exporting suggests that Africa has much to gain from orientating its manufacturing sector towards exporting.

[(1)]Göteborg University, [(2)]University of Oxford, [(3)]The World Bank, [(4)]École des Hautes Études Commerciales, Montréal, [(5)]Free University, Amsterdam, [(6)]University of Ghana, Legon, [(7)]Research Department IMF. The authors form the ISA (Industrial Surveys in Africa) Group, which uses multi-country data sets to analyse the microeconomics of industrial performance in Africa.

This paper draws on work undertaken as part of the Regional Program on Enterprise Development (RPED), organised by the World Bank and funded by the Belgian, British, Canadian, Dutch, French and Swedish governments. Support of the British, Dutch, French and Swedish governments for workshops of the group is gratefully acknowledged. The use of the data and the responsibility views expressed are those of the authors. We would like to thank Alberto Isgut and two anonymous referees for several constructive comments on an earlier version of the paper.

Address for correspondence: Måns Söderbom, Centre for the Study of African Economies, Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ. Måns Söderbom is the UNIDO Research Fellow at the Centre for the Study of African Economies, University of Oxford.
E-mail: Mans.Soderbom@economics.oxford.ac.uk

I. Introduction

It is often argued that trade liberalisation and an export-oriented strategy increase firm-level efficiency (Krugman, 1987; Rodrik, 1988, 1991; Grossman and Helpman, 1991). However, although this is supported by some evidence describing the association between exporting activities and efficiency (Nishimizu and Page, 1982; Haddad, 1993; Harrison, 1994; Aw and Hwang, 1995), there is as yet little systematic evidence that exporting *causes* efficiency gains. Indeed, causality may run in the other direction: efficient firms may self-select into the export market.

One of the first studies that analysed the causal relationship between exporting and productivity at the firm-level was on the U.S. economy (Bernard and Jensen, 1995; see also Bernard and Jensen, 1999a, 1999b). These authors find little evidence of any learning-by-exporting effect. There are now a number of studies examining the link between exporting and productivity on countries other than the USA, see Tybout and Westbrook (1995) on Mexico; Clerides et al. (1998; henceforth CLT) on Mexico, Colombia and Morocco; Kraay (1999) on China; Aw et al. (2000) on the Republic of Korea and Taiwan; Söderbom and Teal (2000) on Ghana; Isgut (2001) on Colombia; and Fafchamps et al. (2002) on Morocco. On balance, there is little evidence in these studies that firms improve their efficiency as a result of a learning-by-exporting process. A common conclusion is that efficient firms self-select into the export market.[1]

In this paper we provide cross-country evidence on this issue for sub-Saharan Africa. Our study is based on panel data on manufacturing firms in Cameroon, Ghana, Kenya and Zimbabwe. These countries have had high trade restrictions in the past and are widely regarded as technologically backward (see Bigsten et al., 1999, for a review of the policy environments in the four countries). In such economies the potential gains from exporting are large. Exporting offers the maximum scope for the increased discipline of competition and contact with foreign customers provides the maximum scope for learning opportunities. Thus, if exporting induces efficiency in any environment, it should do so in

[1] Kraay (1999), however, finds some evidence for learning-by-exporting in Chinese industry, mainly among established exporters.

these economies.[2] From a policy perspective, whether or not firms learn from exporting is an important issue. Africa's domestic markets for manufactures are so small that if African countries are to industrialise, it will have to be through exports. At present there is a substantial competitiveness gap, and under learning-by-exporting such a gap can be reduced endogenously through increased international trade.

Several methodological problems arise when attempting to test for, and distinguish between, learning-by-exporting and self-selection effects. Our approach, which is similar to that of CLT, involves simultaneous estimation of a dynamic production function and a dynamic discrete choice model for the decision to export, where we allow for causality running both from efficiency to exporting and from exporting to efficiency. This strategy enables us to control for unobserved heterogeneity in the form of firm specific effects that are correlated across the two equations. In addition we consider an instrumental variables estimator in order to see if our results are robust. A methodological issue to which we devote considerable attention is the manner in which this unobserved heterogeneity should be modelled. We show that alternative models can give radically different results.

The remainder of the paper is organized as follows. Section II presents our empirical framework and the econometric methods. Section III provides an overview of the data. Section IV reports econometric results analysing the relationship between firm-level efficiency and export history. Section V concludes.

[2] There is, however, a literature which argues that firm productivity in Africa can only be increased by interventions aimed at improving skills and the technical capacity of the firms to absorb new technology (Lall, 1990; Pack, 1993). These authors would argue that such improvements are necessary before firms can become internationally competitive.

II. Empirical Framework

We analyse the link between exporting and efficiency using a production function approach. Our baseline production function is taken to be dynamic Cobb-Douglas, modelling output as a function of capital, labour and intermediate inputs:

$$y_{it} = l y_{i,t-1} + (1 - l)\{b_n n_{it} + b_k k_{it} + b_m m_{it} + b_e e_{it}\} + \log A_{it} + h_{it},$$

where y_{it} is log output, n_{it} is log employment, k_{it} is log capital stock, m_{it} is log raw material, e_{it} is log indirect costs (e.g. electricity, water, transport etc.), A_{it} is total factor productivity, or efficiency, l and b denote parameters to be estimated, h_{it} is a residual, assumed serially uncorrelated, that captures efficiency shocks and i = 1,2,...,N and t = 1,2,...,T are firm and time indices, respectively.[3] In the empirical analysis we consider the effects of allowing for a more flexible specification than Cobb-Douglas, as well as modelling value-added rather than gross output.[4]

Based on the learning-by-exporting idea, we hypothesise that A_{it} depends on exporting and, as learning is unlikely to be instantaneous, that this effect operates with a one-period lag.[5] We allow for heterogeneity in A_{it} by including dummy variables for country, industry, time and firm status (ownership), summarised by the vector c_{it}, and for unobserved heterogeneity in the form of firm specific effects, denoted m_i. We hence write A_{it} in logarithmic form as

$$\log A_{it} = d \cdot exports_{i,t-1} + c_{it} + m_i,$$

---

[3] The dynamic formulation allows for a partial adjustment process of output, reflecting the possibility that, whenever factors of production are changed, it may take time for output to reach its new long-run level. The inclusion of a lagged dependent variable also makes serial correlation of the residual less likely (Nickell, 1996).

[4] Value-added production functions appear to be more common in the literature, however research by Basu and Fernald (1995) shows that adopting a value-added production function can yield misleading results if there is imperfect competition or increasing returns to scale.

[5] Our sample, described in Section III, consists of three waves of panel data. Given the presence of unobserved firm effects, we cannot allow for a longer lag structure than one period. Because the production function contains a lagged dependent variable, entry into, or exit from, the export market will nevertheless have a dynamic effect in that efficiency will be affected for several subsequent time periods. Hence if, for instance, a firm exits at time t it will experience a gradual decline in its efficiency. In the initial periods after exiting its efficiency will be higher than that of an otherwise identical firm that has never exported, but in the long run the efficiency levels of the two firms will converge to the same level.

where *exports* is a dummy variable equal to one if there is some exporting and zero if there is not. Substituting this expression into the production function yields

$$y_{it} = l y_{i,t-1} + (1 - l)\{b_n n_{it} + b_k k_{it} + b_m m_{it} + b_e e_{it}\} + d \cdot exports_{i,t-1} + c_{it} + m_i + h_{it}, \qquad (1)$$

which forms the basis for our econometric test for learning effects that are due to exporting.

A simple empirical approach would be to estimate (1) using for instance OLS or the standard panel GLS ("random effects") estimator. Unfortunately this approach is likely to yield misleading results if exports and productivity are correlated for reasons other than causality running from exports to efficiency. This is emphasised by CLT, arguing that the positive association between export status and productivity can be due to the self-selection of the relatively more efficient plants into foreign markets, rather than learning. In the econometric analysis CLT deal with this problem by formulating a model for export participation in which they control for unobserved firm effects that are potentially correlated with the unobserved firm effects in the productivity equation. We use a similar approach in this paper.

We assume export participation to depend on previous export participation, firm size, labour productivity, capital intensity and a vector of control variables d_{it}. Previous export participation is included in the model to control for fixed costs associated with entering the export market (Roberts and Tybout 1997). Similarly firm size, measured here as the natural logarithm of employment, has a fixed costs interpretation in that exporting typically is associated with costs too large for small firms to incur; for instance, it may be necessary for the exporting firm to set up a marketing department to investigate marketing channels, meet export orders etc. Labour productivity and capital intensity are included in the model to capture a potential self-selection process noted by CLT, by which certain firms choose to export because they are relatively efficient. Given that we control for capital intensity, the coefficient on the labour productivity term is interpretable an efficiency effect.[6] Thus we represent efficiency by

[6] To see this, notice that if the production function is two-factor constant-returns Cobb-Douglas, the residual is equal to $(y-n)-\left(\widetilde{b}_k(k-n)+\text{controls}\right)$, where $\widetilde{b}_k$ denotes the (long-run) capital coefficient. Hence, given that $(k-n)$ is included in the model, the coefficient on $(y-n)$ is interpretable as the effect of the production function residual on exporting.

observables rather than relying on a two-stage procedure where efficiency initially is estimated and then used as a regressor in the export equation.

Because our exports variable is binary we employ a latent variable formulation and, taking the above into account, write the exports equation as

$$\begin{aligned} exports^{*}_{it} = g \cdot exports_{i,t\text{-}1} + q_n \cdot n_{it} + q_y \left(y_{i,t-1} - n_{i,t-1}\right) + q_k \cdot \left(k_{i,t-1} - n_{i,t-1}\right) + \\ d_{it} + y_i + w_{it}, \end{aligned} \qquad (2)$$

where we observe $exports_{is} = 1$ if $exports^{*}_{is} \geq 0$, otherwise zero. Here, g and q denote parameters to be estimated, y_i is an unobserved firm specific time invariant effect affecting the decision to export and w_{it} is a homoskedastic, serially uncorrelated and normally distributed residual whose variance we normalise to one. These assumptions about the residual imply that we can estimate the parameters of interest using a dynamic probit model. We assume that self-selection into exporting operates with a one-period lag, reflected in (2) by the t-1 subscripts on labour productivity and capital intensity. The coefficient q_y thus represents the self-selection effect.

Estimation of (1)-(2) will shed light on, *inter alia* i) if there is support for the learning-by-exporting hypothesis, i.e. that firms improve efficiency as a result of exporting (in which case d would be positive); ii) if there is support for self-selection-into-exporting, i.e. that efficient firms become exporters (in which case q_y would be positive); iii) if there are fixed costs associated with exporting, so that firms tend to continue exporting once they have entered the international market (in which case g would be positive; Roberts and Tybout, 1997). Because the models contain lagged dependent variables it is crucial to control for heterogeneity between firms or we would expect the estimates to be upward biased, reflecting ‘spurious’ state dependence (Heckman, 1981a, 1981b). While the vectors of control variables c_{it} and d_{it} control for heterogeneity in certain observed variables, presence of unobserved heterogeneity in the form of the firm specific effects m_i and y_i presents us with some econometric problems. These are discussed next.

Estimation

In estimating (1)-(2) we mainly rely on maximum likelihood (ML) methods, although we also consider a generalised methods of moments (GMM; Hansen, 1982) estimator. We use three distinct ML models, all of which assume that m_i and y_i can be modelled by means of a random effects approach, and which only differ in what is assumed about the error structure. Equations (1)-(2) contain four random terms, namely m_i, h_{it}, y_i and w_{it}. In our most general ML model we assume that

$$(m_i, y_i, h_{it}, w_{it}) \sim G(z, \Omega), \tag{3}$$

where G is some distribution function, $z = (z_m, z_y, 0, 0)$, and

$$\Omega = \left|\begin{array}{cccc} s_m^2 & & & \\ s_{my} & s_y^2 & & \\ 0 & 0 & s_h^2 & \\ 0 & 0 & s_{hw} & s_w^2 \end{array}\right| = \left|\begin{array}{cccc} s_m^2 & & & \\ r_{my} s_m s_y & s_y^2 & & \\ 0 & 0 & s_h^2 & \\ 0 & 0 & r_{hw} s_h & 1 \end{array}\right|, \tag{4}$$

where r_{my} and r_{hw} denote the correlation between m and y, and between h and w, respectively. Thus the transitory errors h_{it} and w_{it} are taken to be uncorrelated with the permanent effects m_i and y_i, an assumption we make for computational reasons. Throughout the analysis we assume that h_{it} and w_{it} follow a bivariate normal distribution.

Our simplest model imposes the restriction $s_m^2 = s_y^2 = r_{my} = 0$, which amounts to assuming that there is no unobserved heterogeneity in the form of firm specific effects. In this special case the likelihood function can be written ignoring the panel nature of the data altogether. While this model is straightforward to estimate, the presence of dynamic terms in the regression means that consistency of the estimates hinges crucially on the absence of unobserved heterogeneity. Even though the model thus is rather restrictive, it is useful as a benchmark. The likelihood function is shown in the Appendix, Part A.

The second model is similar to that used by CLT in assuming that m_i and y_i follow a bivariate normal distribution.[7] In this case the likelihood function involves multiple integrals which makes computation rather more difficult than for our first model. Following CLT we deal with this by integrating out the random effects m_i and y_i using a bivariate Gaussian quadrature. Details on this procedure, and the likelihood function, are given in Part B of the Appendix. In the remainder of the paper we refer to this model as the CLT model.

Although the CLT model is attractive in that it allows for unobserved firm effects that are correlated across the two equations, the distributional assumptions about the error structure are potentially restrictive. In our third ML model we relax the assumption that m_i and y_i are normally distributed, and follow Heckman and Singer (1984) in adopting a non-parametric strategy for characterising the distribution of the random effects. Specifically, we assume that the bivariate distribution of m_i and y_i can be approximated by a discrete multinomial distribution with *Q* x *R* points of support:

$$\Pr\left(m = m_q, y = y_r\right) = P_{qr}\ , \qquad q = 1,2,\ldots,Q\ ; \quad r = 1,2,\ldots,R\ ; \quad \sum_{q=1}^{Q}\sum_{r=1}^{R} P_{qr} = 1,$$

where the m_q, y_r and P_{qr} are parameters to be estimated.[8] Hence, the estimated support points determine where the observations are positioned, and P_{qr} indicate the proportion of the observations found at each particular point. This model is flexible and several restrictions inherent in the CLT model (e.g. symmetric distribution of heterogeneity) are avoided.[9] In estimating the model, one important issue refers to the number of support points, *Q* and *R*. In fact, there are no well-established criteria for determining the number of support points in models like these (see e.g. Heckman and Walker, 1990), so we follow standard practice and increase *Q* and *R* until there are only marginal improvements in the log

---

[7] A similar specification has also been used by Keane et al. (1988) in their analysis of real wages over the business cycle.

[8] The multinomial approach to characterising the distribution of heterogeneity has been used in various microeconometric analyses of, for instance, dynamic discrete choice (Moon and Stotsky, 1993; Blau and Gilleskie, 2001), duration data (Blau, 1994; Ham and LaLonde, 1996), and count data (Deb and Trivedi, 1997).

[9] Monte Carlo evidence indicates that this approach compares favourably to standard ML correctly assuming a normal distribution, and that it performs much better than ML incorrectly assuming normality

likelihood value. Usually, the number of support points is small; indeed, for Q = 1, R = 1 unobserved heterogeneity is absent and the production function and the exports equation are independent, implying that our first model discussed above may be used to estimate the parameters of interest. The likelihood function is given in the Appendix, Part C. In the remainder of the paper we refer to this model as the NPML model.

In forming the likelihood underlying the CLT and NPML models, we have to recognise the presence of lagged dependent variables among the explanatory variables. This creates an initial conditions problem in that $exports_{i0}$ and y_{i0} will be correlated with the firm specific effects if $exports_{i0}$ and y_{i0} have been determined by the same model as that determining productivity and exports from t = 1 and onwards.[10] Neglecting the initial conditions problem leads to inconsistent parameter estimates unless T is large. Following Heckman (1981a, 1981b) we approach this problem by specifying models for the initial conditions $exports_{i0}$ and y_{i0}, allowing these variables to depend on the random effects m_i and y_i by means of a factor loading approach (see Appendix, Part B). The parameters of the initial conditions models are then estimated jointly with the other parameters. CLT use a similar approach.

All ML models discussed above assume that all explanatory variables except $exports_{i,t-1}$ and $\left(y_{i,t-1} - n_{i,t-1}\right)$ are uncorrelated both with the firm specific effects and the transitory errors. This assumption is made for computational reasons, and it is of obvious interest to investigate how strong an assumption this is. While it would be possible to relax the exogeneity assumption for all variables within the ML framework, estimation would have to proceed in one step to avoid a substantial efficiency loss (Mroz, 1999). One-step estimation involves adding additional equations to the system --

---

(Mroz and Guilkey, 1995; Mroz, 1999).

[10] A simple example may illustrate this: Consider a process where y_t depends on y_{t-1} and a random effect z, and define the per-period likelihood contribution as $f(y_t \mid y_{t-1}, z)$. Since z is unobserved we need to integrate over its distribution in order to formulate the likelihood solely in terms of observable variables. If y_0 for some reason is independent of z, the likelihood unconditional of z is simply $\int \Pi_{t=1} f(y_t \mid y_{t-1}, z)\, dG(z)$. In this case there is no difference compared to the static counterpart of the model. However, if y_0 is dependent of z, say because the process begun before the time of the first observation of the sample, the likelihood is equal to $\int \Pi_{t=1} f(y_t \mid y_{t-1}, z)\, h(y_0 \mid z)\, dG(z)$, where $h(y_0 \mid z)$ denotes the marginal density of y_0 given z. Dealing with $h(y_0 \mid z)$

one for each endogenous variable -- and then estimating all equations simultaneously. Unfortunately this becomes increasingly computationally intractable as the number of endogenous variables grows. Instead, we estimate the production function using instrumental variables in a GMM framework, which enables us to test for the validity of the exogeneity assumption and, if rejected, re-estimate the model treating the inputs as endogenous. We estimate the production function both in levels and in first differences and compare the results.

III. Data and Summary Statistics

Our data are for manufacturing firms in four African countries - Cameroon, Kenya, Ghana and Zimbabwe. The data were collected during the period 1992 to 1995 as part of the Regional Program on Enterprise Development (RPED) coordinated by the World Bank. In each country, over a period of three years, a panel of firms in the manufacturing sector was surveyed and information was gathered on a variety of issues, including outputs and resource use. The periods covered by the surveys were as follows: for Kenya, 1992 to 1994; for Ghana, 1991 to 1993; for Zimbabwe, 1992 to 1994; and for Cameroon, 1992/93 to 1994/95. All the countries faced problems in their macroeconomic environments that had a significant impact on manufacturing sector performance. They had all adopted import substitution development policies from independence through the late 1970s. In the mid to late 1980s, they had all introduced 'structural adjustment' programs with the support of the World Bank and other aid organisations, with emphasis on macroeconomic reforms, trade liberalisation and privatisation. The scope and success of these programs varied. For a discussion of policy in the four countries see Bigsten et al. (1999, 2001).

Throughout the paper we use the balanced panel of those firms for which observations exist for all three survey years, because this is the minimum time period necessary to control for unobserved firm effects in the econometric analysis.[11] We correct for changes in exchange rates during the sampling

is the initial conditions problem (see e.g. Hsiao, 1986, pp. 169-172).

[11] If the probability of being included in the sample is correlated with unobservable factors affecting the dependent variables in our regressions, it is possible that this introduces selectivity bias. The most direct

years, given that devaluation would make an exporting firm to appear more productive as the value of its output is valued more highly in terms of local currency, which could induce spurious correlation. To this end we use firm-specific deflators based on export share-weighted averages of the domestic and international prices. In the same manner, inputs are deflated using import shares.[12] Of course, this is unlikely to yield perfect deflators but it is difficult to do better given the data available.

Table 1 shows summary statistics on the estimation sample, where the sample is split by initial export status. Because our regressions include lags we lose one wave of the data, hence only two years of data are being used here. About 29 per cent (85 out of a total of 289 firms) of the firms observed at $t = 0$ are exporters, and within this group of initial exporters the proportion of exporters in the consecutive two years is about 85 per cent. In the group of initial non-exporters, only 8 per cent of the observations record any exporting in the subsequent two years. Hence there is strong persistence in the export data. Further, initial exporters are larger than non-exporters, and exhibit higher labour productivity and higher capital intensity. Sixty-one per cent of the initial exporters are Zimbabwean firms, and 40 per cent of the initial exporters have some foreign ownership. There is no obvious pattern across sectors.

IV. Econometric Analysis

Table 2 reports selected estimates for our baseline specification, using the three ML models discussed in Section II. The production function, taken to be Cobb-Douglas, models gross output. Column [1] shows the results for the simplest model, i.e. where firm effects are ignored altogether. In the production

---

way to remedy this involves specifying a selectivity model. This would be quite complicated given the nature of the models we estimate in Section IV, and we do not attempt to do so in this paper. Nevertheless, if sample selectivity introduces substantial bias we would expect this to be picked up by our instrumental variable estimates (as selectivity bias is a form of omitted variables bias).

[12] Specifically, we begin by constructing firm specific Laspeyres indices as weighted sums of the consumer price index and an index of the nominal exchange rate to the US Dollar. We construct one index for output and one for inputs, using as weights for the output index the percentage of output exported in the initial period, and for inputs the percentage of raw materials imported in the initial period. We then use these indices to deflate output and raw material costs to constant 1992 domestic prices. We deflate physical capital using the nominal exchange rate as physical capital typically is imported and we do not have data enabling us to construct weighted indices. Indirect costs are deflated using the CPI as this input (electricity, water, etc.) typically is not imported. We then convert all monetary variables to PPP adjusted 1992 US Dollars, to ensure

function all inputs in are significant at the five per cent level or better and sum to 0.85, which, given that the coefficient on the lagged dependent variable is 0.16, implies that long-run constant returns to scale can easily be accepted (test not reported). The estimated coefficient on the lagged export variable is equal to 0.07 and significant at the five per cent level, thus suggesting a positive effect of exporting onto efficiency. In the export probit the coefficient on $(y_{t-1} - n_{t-1})$ is positive and significant at the five per cent level. Given that the regression controls for the capital-labour ratio, this is evidence that an increase in efficiency at time t increases the export probability at time t+1, as predicted by the self-selection theory. The coefficient on lagged export is equal to 2.02 and highly significant, indicating strong persistence in the exporting decision.[13] Of course, given that we do not control for time invariant firm effects here, this effect is probably upward biased reflecting 'spurious' state dependence (Heckman, 1981a, 1981b). The coefficient on employment is positive and highly significant. The result that contemporaneous exports is affected by lagged exports and size can be interpreted as evidence for fixed costs (see Section II).

Next consider the effects of allowing for unobserved heterogeneity. Column [2] shows the results of the CLT model in which the firm effects are taken to follow a bivariate normal distribution. The increase in the log likelihood value compared to Column [1] indicates that this model provides a far better fit to the data than the simpler model. Strikingly, there is now no evidence for learning by exporting, as the coefficient on lagged exports is far from significant and the point estimate is even negative. There is unobserved heterogeneity both in the production function and in the export equation, and the estimate of r_{my} indicates that the correlation between m and y is equal to 0.31. This suggests that the positive coefficient on the export variable in Column [1] is upward biased due to the omission of unobserved heterogeneity, consistent with the argument of CLT. Further, in the export equation the coefficient on lagged exports has collapsed to -0.07 and is insignificant. The reason is that the estimate

that the data are comparable across countries.

[13] To translate this into an effect on the probability of exporting, consider a non-exporting firm whose predicted probability of exporting at time t is 0.10. If this firm breaks into the export market at time t then, holding all other characteristics constant, the predicted probability of exporting at time t+1 would increase from 0.10 to about 0.77, clearly a large effect.

of s_y , the standard deviation of the random effect y_i, is very high indeed. This would imply that the observed persistence in the export data documented in Column [1] is entirely due to unobserved time invariant heterogeneity, and not driven by a causal effect of past onto contemporaneous exporting as predicted by the sunk cost model developed by Roberts and Tybout (1997). In the production function all coefficients on the input factors are significant, and the long-run elasticities sum to 1.02. In the export equation the coefficient on labour productivity is positive and close to significant at the ten per cent level, providing only weak evidence for self-selection. The employment coefficient is positive and highly significant.

The CLT results thus provides no evidence in favour of the learning-by-exporting hypothesis. Consider now the effect of relaxing the assumption that m and y are normally distributed. Column [3] reports NPML estimates where the bivariate distribution of m and y is taken to be discrete with 3 x 3 points of support.[14] The resulting log likelihood value is almost 25 units higher than in the CLT model, indicating that the NPML model provides a much better fit to the data. Several results are worth noting. First, the estimated coefficient on lagged exports is equal to 0.07 and significant at the five per cent level. In fact, the point estimate is almost identical to the result shown in Column [1]. Thus we can now reject the null hypothesis that exporting has no effect on efficiency. The lower part of Table 2 shows that the estimated standard deviations of m and y are rather much smaller than in the CLT model, and there is no evidence that m and y are correlated. It is therefore not surprising that some of the coefficients in the production function and the export equation are rather different. Further, in the exports equation the coefficient on lagged exports is now significant and much higher than in the CLT model.[15] Finally, it is noted that the long-run elasticities in the production function sum to 1.02, that in

---

[14] Increasing the number of support points further resulted in a very small increase in the log likelihood value.

[15] The point estimate of the coefficient on lagged exports is equal to 1.075. If a previously non-exporting firm, whose predicted probability of exporting at time t is 0.10 conditional on observables and the unobserved firm effect, breaks into the export market at time t then, holding all other characteristics constant, the predicted probability of exporting at time t+1 would increase from 0.10 to about 0.42. It is noted that this is a much smaller effect than that implied by the results ignoring unobserved heterogeneity, see footnote 12. This is an example of how ignoring unboserved time invariant heterogeneity in dynamic models leads to ‘spurious state dependence’, Heckman (1981a, 1981b).

the export equation the coefficient on labour productivity is positive and significant at the ten per cent level and that the employment coefficient is positive and highly significant.

Allowing for a more flexible form of unobserved heterogeneity than that based on the bivariate normal distribution has led to radically different estimates of the associated moments, which has far-reaching implications for the estimates of the coefficients of interest. Table 3 shows the NPML estimate of the joint probability distribution of m and y. Clearly the distribution is quite asymmetrical, which suggests that joint normality will be a restrictive assumption for these data. We now probe the data further, in order to investigate if the learning-by-exporting result obtained in Table 2, Column [3], is robust to alternative specifications. We begin by considering a more flexible functional form for the production function. One flexible form that has been used extensively in studies estimating cost and production functions is the second-order transcendental logarithmic ('translog') production function (Christensen et al., 1971; Berndt and Christensen, 1972). This is a generalisation of the Cobb-Douglas model that includes squared and interacted terms of the factor inputs (in natural logarithms), in addition to the levels terms. Output elasticities hence vary with the levels of the inputs, and to facilitate interpretation of the results we therefore report elasticities evaluated at sample means of the regressors. The non-linear form of the translog model also implies that the regularity conditions of the production function, notably monotonicity and quasi-concavity, will have to be investigated at each data point.[16]

In Table 4, Columns [1]-[2] we report results where the production function is assumed to be second-order translog, for the CLT and the NPML specifications. In both cases there is a significant increase in the log likelihood value, suggesting that the translog specification provides a better approximation of the technology than the Cobb-Douglas model.[17] Evaluated at sample means, the estimated elasticities of the inputs are nevertheless similar to the Cobb-Douglas coefficients. As for the

---

[16] Monotonicity requires that each input has a positive marginal product, and quasi-concavity requires that the bordered Hessian matrix of first and second partial derivatives of the production function are negative semi-definite.

[17] It is noted however that the translog model complies relatively poorly with monotonicity and quasi-concavity: about 70 per cent of the observations comply with monotonicity, while only about 35 per cent of the observations are consistent with quasi-concavity. It is possible that this is driven by imprecise estimates of the elasticities.

effect of exporting on efficiency, the results of the CLT and NPML models are very similar to their Cobb-Douglas counterparts. In the CLT model the coefficient on exporting is negative and insignificant, while in the NPML model the coefficient is about 0.07 and statistically significant at the five per cent level. Again, the log likelihood value of the NPML model is much higher than that of the CLT model suggesting that the former provides a better fit to the data. Thus, while the translog specification may seem preferable to the Cobb-Douglas model, the effect of exporting on efficiency appears not to be sensitive to the functional form of the production function. In the export equation there is evidence for a strong size effect in both specifications, and for self-selection and persistence in the NPML model. For both models the estimated distribution of the random effects is similar to the results in Table 3.

Thus far the production function has modelled gross output as a function of capital, labour and intermediate inputs. In Table 4, Columns [3]-[5] we report production functions that model value-added, defined as the value of output minus the value of raw materials and indirect costs, with capital and labour as the factor inputs. As expected, the production function coefficients are much larger in magnitude than in the gross output production function.[18] The model without firm effects yields a positive and significant coefficient on the exports variable, while the CLT model again yields a negative and insignificant coefficient, providing no evidence for learning. In estimating the NPML model the variance of m systematically tended to zero despite using more than 100 different vectors of start values, and we therefore imposed zero variance in this model. The NPML parameter estimates of the production function are similar in magnitude to those shown in Column [3]. Most notably, the coefficient on the export variable is positive and significant at the five per cent level. Thus the pattern is the same as that documented earlier, in that imposing bivariate normality on these data dramatically affects the coefficient on the export dummy. Again the NPML model yields the highest log likelihood

[18] To see this, assume for simplicity that the cost of raw materials and indirect inputs is a constant fraction of output. In this case the long-run coefficients on factor inputs in the output model are scaled up by the inverse of one minus the sum of the long-run coefficients on the intermediate inputs, to yield value-added equation coefficients.

value. In the export equation there is a strong size effect in all specifications, while the evidence for a self-selection effect is weak except in the model without firm effects.

The evidence thus seems quite clear that assuming the random effects to follow a bivariate normal distribution is an incorrect assumption for these data, and that imposing bivariate normality has a considerable effect on some of the parameter estimates. Most notably, under bivariate normality there is no significant exporting effect on productivity. Why is the normality assumption problematic in the current application? In the univariate case there is a fairly large literature discussing parametric assumptions regarding the distribution of unobserved heterogeneity. We are unaware of any paper discussing this issue for bivariate distributions.[19] Further inspection of Tables 2 and 4 gives us some clues of the nature of the problem. The CLT exports coefficient in the production function is imprecisely estimated. In Table 2 its 95 per cent confidence interval ranges between –0.091 and 0.089. Further, the low CLT estimate of the exports coefficient is accompanied by a relatively high estimate of r_{my}, measuring the correlation of the two time invariant random effects. A log-likelihood ratio test reveals, however, that this coefficient is not significantly different from zero. Estimating the CLT model in Table 2 imposing $r_{my} = 0$ yields a point estimate of the exports coefficient equal to 0.06, which is very similar to the NPML model. Hence under bivariate normality we obtain something similar to an identification problem, where it is difficult to distinguish between a causal effect and time invariant heterogeneity. Investigating whether this is a general result for models of this kind is left for future research. What seems clear is that a more flexible characterisation of the distribution of the random effects greatly improves our ability to pin down the parameters of interest in the model.

Finally we estimate the production function using an instrumental variables approach to assess if the above results are biased by simultaneity. The first column of Table 5 shows two-step GMM estimates of the Cobb-Douglas output production function in levels, where the *t*-statistics are based on

---

[19] It has long been known in the duration literature, for instance, that models based on parametric assumptions about the hazard function and the heterogeneity distribution can lead to seriously biased results (Heckman and Singer, 1984).

robust, finite sample corrected standard errors (see Windmeijer, 2000).[20] We include in our instrument set, specified in its entirety in the notes to Table 5, contemporaneous values of the factor inputs as this will shed light on whether the assumption that these are uncorrelated with the residual is valid. The Sargan-Hansen *J*-statistic, equal to the value of the criterion function evaluated at the GMM estimates, implies that we cannot reject the hypothesis that the overidentifying restrictions are valid. Hence there is no evidence that including contemporaneous values of the regressors in the instrument set results in a misspecification. The coefficient on the lagged dependent variable is lower than in the model without firm effects shown in Table 2, which is to be expected if there are unobserved firm specific effects. The coefficients on labour and intermediate inputs are positive and significant at the one per cent level, while the capital coefficient is positive and significant at the ten per cent level. The point estimate of the export coefficient is equal to 0.06, hence very similar to the NPML results, and significant at the ten per cent level.

Next we take first differences of the production function which removes the firm specific effect. To deal with the bias of the coefficient on the lagged dependent variable identified by Nickell (1981) we use output lagged two periods as an instrument. Additional instruments are lagged and contemporaneous values of the inputs. Results are shown in Column [2]. Again, the *J*-statistic indicates that exogeneity of the inputs is not an overly restrictive assumption as we can accept the validity of the overidentifying restrictions. Some of the point estimates are quite different compared to the levels specification, however the t-values are lower than previously reflecting larger standard errors. This is not surprising, given that the first differencing procedure reduces the variation in the explanatory variables. While the coefficient on exports is no longer significant, the point estimate is larger than previously. Hence this lack of significance is probably due to this estimator being less efficient than the

[20] It is well known that the asymptotic standard errors in two-step GMM estimators can be severely downward biased in finite samples (e.g. Arellano and Bond, 1991). As a consequence, researchers often draw inference based on one-step GMM estimators, which are less efficient than the two-step estimators. However, Windmeijer (2000) shows how the asymptotic two-step standard errors can be corrected when the sample size is finite. Monte Carlo evidence reported by Bond and Windmeijer (2001) indicates that this procedure yields a much more reliable basis for inference than relying on the asymptotic standard errors.

levels estimator and the ML models. Thus it there is little evidence that the assumption of strict exogeneity of the factor inputs is too restrictive.[21]

V. Conclusion

In this paper, we have examined two not incompatible explanations for the positive association between export-participation status and productivity: self-selection of the relatively more efficient plants into exporting, and learning by exporting, using panel data on manufacturing firms in four African countries. Our preferred estimates show that, consistent with the learning-by-exporting hypothesis, exporting impacts positively on productivity. This result is not sensitive to the functional form of the production function, and neither is there any evidence that neglected simultaneity is driving the result. There is some evidence for self-selection into the export market, thus suggesting that causality runs both from exporting to efficiency and from efficiency to exporting. The evidence also indicates that that past exporters are more likely to remain active in the export market, consistent with the presence of sunk cost of breaking into the foreign markets (Roberts and Tybout, 1997).

There is strong evidence for unobserved heterogeneity in the data, only detectable with panel data. It is however quite clear that assuming the random effects to follow a bivariate normal distribution is an incorrect assumption. Using a more flexible specification yields a highly asymmetrical distribution of the firm effects, which is inconsistent with normality. Further, imposing bivariate normality on the data has a considerable effect on some of the parameter estimates. Most notably, under bivariate normality there is no significant exporting effect on productivity. It is noted that CLT obtain a similar result, leading the authors to conclude that '...the association between exporting and efficiency is most plausibly explained as low-cost producers choosing to become exporters.' (CLT, p. 942). Had we confined ourselves to the CLT model, our conclusion would probably have been the

[21] A similar result has been obtained by Söderbom and Teal (2002), estimating production functions using seven years of panel data on Ghanaian manufacturing firms, so this finding is not an artefact of our sample.

same.

From a policy perspective, the result that there is learning-by-exporting is an important one. Africa’s domestic markets for manufactures are so small that if African countries are to industrialise, it will have to be through exports. Our results provide strong support for the view that learning-by-exporting is possible in Africa. If this is so, Africa has much to gain from orientating its manufacturing sector towards exporting.

References

Arellano, M. and S. Bond, 1991, 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,' *Review of Economic Studies* 58, pp. 277-297.

Aw, B., S. Chung and M. J. Roberts, 2000, 'Productivity and Turnover in the Export Market: Micro-level Evidence from the Republic of Korea and Taiwan (China),' *World Bank Economic Review* 14, pp. 65-90.

Aw, B. Y. and A. R. Hwang, 1995, 'Productivity and the Export Market: A Firm-Level Analysis,' *Journal of Development Economics* 47, pp. 313-332.

Basu, S., and J. G. Fernald, 1995, 'Are Apparent Productive Spillovers a Figment of Specification Error?' *Journal of Monetary Economics* 36, pp. 165-188.

Bernard, A. B. and J. B. Jensen, 1995, 'Exporters, Jobs and Wages in US Manufacturing 1976-87,' *Brookings Papers on Economic Activity Microeconomics*, pp. 67-119.

___________ 1999a, 'Exporting and Productivity,' mimeo. Yale University.

___________ 1999b, 'Exceptional Exporter Performance: Cause, Effect, or Both?' *Journal of International Economics* 47, pp. 1-25.

Berndt, E. and L. Christensen, 1972, 'The Translog Function and the Substitution of Equipment, Structures, and Labor in U.S. Manufacturing, 1929-1968,' *Journal of Econometrics* 1, pp. 81-114.

Bigsten, A., P. Collier, S. Dercon, M. Fafchamps, B. Gauthier, J. W. Gunning, J. Haraburema, A. Oduro, R. Oostendorp, C. Pattillo, M. Söderbom, F. Teal and A. Zeufack, 1999, 'Exports of African Manufactures: Macro Policy and Firm Behaviour', *The Journal of International Trade and Economic Development* 8, pp. 53-71.

Bigsten, A., P. Collier, S. Dercon, M. Fafchamps, B. Gauthier, J. W. Gunning, A. Oduro, R. Oostendorp, C. Pattillo, M. Söderbom, F. Teal and A. Zeufack, 2001, 'Exporting and Efficiency in African Manufacturing,' pp. 111-121 in *Policies to Promote Competitiveness in Manufacturing in Sub-Saharan Africa*, edited by A. K. Fosu, S. M. Nsouli and A. Varoudakis. Paris: IMF, AERC and Development Centre of the OECD.

Blau, D. M., 1994, 'Labor Force Dynamics of Older Men,' *Econometrica* 62, pp. 117-156.

Blau, D. M. and D. B. Gilleskie, 2001, 'Retiree Health Insurance and the Labor Force Behavior of Older Men in the 1990s,' *Review of Economics and Statistics* 83, pp. 64-80.

Bond, S. and F. Windmeijer, 2001, 'Finite Sample Inference for GMM Estimators in Linear Panel Data Models,' mimeo. London: The Institute for Fiscal Studies.

Christensen, L. R., D. W. Jorgensen and L. J. Lau, 1971, 'Transcendental Logarithmic Production Frontiers,' *Review of Economics and Statistics* 55, pp. 28-45.

Clerides S., S. Lach and J. Tybout, 1998, 'Is Learning by Exporting Important? Micro-Dynamic Evidence From Colombia, Mexico and Morocco,' *Quarterly Journal of Economics* 113, pp. 903-948.

Deb P. and P. K. Trivedi, 1997, 'Demand for Medical Care by the Elderly: a Finite Mixture Approach,' *Journal of Applied Econometrics* 12, pp. 313-336.

Fafchamps, M., S. El Hamine and A. Zeufack, 2002, 'Learning to Export: Evidence from Moroccan Manufacturing,' CSAE Working Paper WPS/2002-02. Oxford: Centre for the Study of African Economies, Oxford University.

Grossman, G. and E. Helpman, 1991, *Innovation and growth in the Global Economy.* Cambridge, Massachussets: The MIT Press.

Haddad, M., 1993, 'How Trade Liberalization Affected Productivity in Morocco,' Policy Research Working Paper no 1096. Washington D.C.: The World Bank.

Ham, J. C. and R. J. LaLonde, 1996, 'The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training,' *Econometrica* 64, pp. 175-205.

Hansen, L. P., 1982, 'Large Sample Properties of Generalized Method of Moments Estimators,' *Econometrica* 50, pp. 1029-1054.

Harrison, A., 1994, 'Productivity, Imperfect Competition, and Trade Reform: Theory and Evidence,' *Journal of International Economics* 36, pp. 53-73.

Heckman, J. J., 1981a, ‘Statistical Models for Discrete Panel Data,’ in *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. F. Manski and D. McFadden. Cambridge, Massachusetts: The MIT Press.

___________ 1981b, ‘The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process,’ in *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. F. Manski and D. McFadden. Cambridge, Massachusetts: The MIT Press.

Heckman, J. J. and B. Singer, 1984, ‘A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data,’ *Econometrica* 52, pp. 271-320.

Heckman, J. J. and J. Walker, 1990, ‘The Relationship Between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data,’ *Econometrica* 58, pp. 1411-1441.

Hsiao, C., 1986, *Analysis of Panel Data*. Melbourne: Cambridge University Press.

Isgut, A. E., 2001, ‘What’s Different about Exporters? Evidence from Colombian Manufacturing,’ *Journal of Development Studies* 37, pp. 57-82.

Judd, K. L., 1998, *Numerical Methods in Economics*. Cambridge, Massachusetts; London, England: The MIT Press.

Keane, M, R. Moffit and D. Runkle, 1988, ‘Real Wages Over the Business Cycle: Estimating the Impact of Heterogeneity with Micro Data,’ *Journal of Political Economy* 96, pp. 1232-1265.

Kraay, A., 1999, ‘Exportations et performances economiques: Etude d'un panel d'entreprises chinoises, *Revue d'Economie du Developpement* 0: 183-207.

Krugman, P., 1987, ‘The Narrow Moving Bank, The Dutch Disease and the Competitive Consequences of Mrs Thatcher: Notes on Trade in the Presence of Dynamic Scale Economies,’ *Journal of Development Economics* 27, pp. 41-55.

Lall, S., 1990, *Building Industrial Competitiveness in Developing Countries*. Paris: OECD.

Moon, C. and J. G. Stotsky, 1993, ‘The Effect of Rent Control on Housing Quality Change: A Longitudinal Analysis,’ *Journal of Political Economy* 101, pp. 1114-1148.

Mroz, T.A., 1999, 'Discrete Factor Approximations in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on a Continuous Outcome', *Journal of Econometrics* 92, pp. 233-274.

Mroz, T. A. and D. Guilky, 1995, 'Discrete Factor Approximations For Use in Simultaneous Equation Models with Both Continuous and Discrete Endogenous Variables,' mimeo. University of North Carolina at Chapel Hill.

Nickell, S. J., 1981, 'Biases in Dynamic Models with Fixed Effects,' *Econometrica* 49, pp. 1399-1416.

___________ 1996, 'Competition and Corporate Performance,' *Journal of Political Economy* 104, pp. 724-746.

Nishimizu, M. and J. M. Page Jr., 1982, 'Total Factor Productivity Growth, Technological Progress and Technical Efficiency Change: Dimensions of Productivity Change in Yugoslavia 1967-78,' *Economic Journal* 92, pp. 920-36.

Pack, H., 1993, 'Productivity and Industrial Development in Sub-Saharan Africa,' *World Development* 21, pp. 152-176

Roberts, M. and J. Tybout, 1997, 'The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs,' *American Economic Review* 87, pp. 545-64.

Rodrik, D., 1988, 'Imperfect Competition, Scale Economies and Trade Policy in Developing Countries,' in *Trade Policy Issues and Empirical Analysis*, edited by R.E. Baldwin. Chicago: University of Chicago Press.

___________ 1991, 'Closing the Technology Gap: Does Trade Liberalization Really Help?' in *Trade Policy, Industrialization and Development: New Perspectives*, edited by G. Helleiner. Oxford: Clarendon Press.

Söderbom, M. and F. Teal, 2000, 'Skills, Investment and Exports from Manufacturing Firms in Africa,' *Journal of Development Studies* 37, pp. 13-43.

___________, 2002, 'Size and Efficiency in African Manufacturing Firms: Evidence from Firm-Level Panel Data,' mimeo. Oxford: Centre for the Study of African Economies, Department of Economics, University of Oxford.

Tybout, J. R. and M. D. Westbrook, 1995, 'Trade Liberalization and the Dimensions of Efficiency Change in Mexican Manufacturing Industries,' *Journal of International Economics* 39, pp. 53-78.

Windmeijer, F., 2000, 'A Finite Sample Correction for the Variance of Linear Two-step GMM Estimators,' IFS Working Paper 00/19. London: The Institute for Fiscal Studies.

TABLE 1

SUMMARY STATISTICS, BY INITIAL EXPORT STATUS

	Initial Exports = 0 (Number of firms: 204)			Initial Exports = 1 (Number of firms: 85)		
	Mean	p50	Std. Dev.	Mean	p50	Std. Dev.
VARIABLE IN *t* = 1,2						
Exports	0.08			0.85		
Employment	51.44	20.00	104.54	342.07	171.00	492.19
Ln Employment	3.04	3.00	1.29	5.12	5.14	1.25
Ln Value-Added / Employee	8.16	8.32	1.34	9.37	9.35	0.99
Ln Output / Employee	9.31	9.50	1.23	10.37	10.22	0.85
Ln Physical Capital / Employee	7.93	8.21	1.84	9.42	9.43	1.13
Cameroon	0.35			0.14		
Ghana	0.26			0.08		
Kenya	0.20			0.16		
Zimbabwe	0.19			0.61		
Food	0.26			0.19		
Textile	0.26			0.26		
Metal	0.23			0.28		
Wood	0.25			0.27		
Any Foreign Ownership	0.18			0.40		
Any State Ownership	0.03			0.08		

Note: Variables for which p50 and Std. Dev. are not reported are dummy variables.

TABLE 2

SELECTED MAXIMUM LIKELIHOOD ESTIMATES:

COBB-DOUGLAS OUTPUT PRODUCTION FUNCTION AND EXPORT PROBIT

	[1] No firm effects	[2] Bivariate normal firm effects (CLT)	[3] Non-parametric bivariate firm effects (NPML)
THE PRODUCTION FUNCTION			
y_{t-1}	0.162	0.097	0.118
	(8.684)**	(5.122)**	(6.284)**
$export_{t-1}$	0.068	-0.001	0.067
	(2.093)*	(0.020)	(2.142)*
k_t	0.021	0.027	0.034
	(2.114)*	(2.547)*	(3.459)**
n_t	0.103	0.143	0.112
	(5.539)**	(6.606)**	(6.001)**
m_t	0.627	0.666	0.668
	(37.216)**	(41.270)**	(39.940)**
e_t	0.094	0.089	0.083
	(6.611)**	(6.226)**	(6.078)**
THE EXPORT EQUATION			
$(y_{t-1} - n_{t-1})$	0.228	0.320	0.259
	(1.984)*	(1.518)	(1.828)$^{+}$
$export_{t-1}$	2.021	-0.067	1.075
	(10.816)**	(0.194)	(3.017)**
$(k_{t-1} - n_{t-1})$	0.047	-0.030	0.021
	(0.635)	(0.259)	(0.238)
n_{t-1}	0.273	1.918	0.597
	(3.420)**	(4.612)**	(3.297)**
s_h	0.269	0.224	0.242
s_m		0.160	0.126
s_y		2.521	0.807
r_{hw}	0.052	-0.211	0.024
r_{my}		0.311	-0.019
Log likelihood value	-378.13	-357.42	-333.07
Number of firms	289	289	289

Note: All regressions include dummy variables for country, industry, ownership and time. The numbers in ( ) are *t*-statistics based on asymptotic standard errors. Significance at the one per cent, five per cent and ten per cent level is indicated by *, ** and $^{+}$ respectively.

TABLE 3

THE NPML ESTIMATE OF THE PROBABILITY DISTRIBUTION OF m AND y

		y			
		y_1: -0.577	y_2: 0.879	y_3: 2.189	$f_y(y)$
	m_1: -0.052	0.494	0.348	0.000	*0.842*
m	m_2: 0.266	0.133	0.000	0.022	*0.155*
	m_3: 0.797	0.000	0.000	0.003	*0.003*
	$f_m(m)$	*0.627*	*0.348*	*0.026*	

Note: The table shows the estimated probability distribution based on the model reported in Table 2, Column [3]. The positions of the random effects are indicated by $m_1,...,m_3$, $y_1,...,y_3$. Four of the estimated joint probabilities tended to zero when estimated freely. To obtain a non-singular Hessian we consequently imposed zero values on these probabilities. $f_m(m)$ and $f_y(y)$ indicate the marginal probabilities.

TABLE 4

SELECTED PRODUCTION FUNCTION AND EXPORT EQUATION

MAXIMUM LIKELIHOOD ESTIMATES

	Production Function: Gross Output, Translog[$]		Production Function: Value-Added, Cobb-Douglas		
	[1] CLT	[2] NPML	[3] No firm effects	[4] CLT	[5] NPML
THE PRODUCTION FUNCTION					
y_{t-1}	0.087	0.104	0.482	0.287	0.479
	(4.853)**	(5.940)**	(13.659)**	(3.938)**	(13.567)**
$export_{t-1}$	-0.037	0.068	0.232	-0.071	0.234
	(0.740)	(2.285)*	(2.261)*	(0.452)	(2.283)*
k_t	0.025	0.034	0.140	0.193	0.138
	(2.382)*	(3.553)**	(4.862)**	(5.637)**	(4.794)**
n_t	0.144	0.114	0.454	0.656	0.460
	(6.609)**	(6.150)**	(8.515)**	(8.013)**	(8.632)**
m_t	0.664	0.665			
	(40.022)**	(38.648)**			
e_t	0.115	0.107			
	(7.061)**	(8.179)**			
THE EXPORT EQUATION					
$(y_{t-1} - n_{t-1})$	0.232	0.256			
	(1.333)	(1.813)[+]			
(log value-added $_{t-1}$ - n_{t-1})			0.215	0.079	0.172
			(2.075)*	(0.479)	(1.156)
$export_{t-1}$	0.384	1.106	2.010	0.504	0.533
	(1.054)	(2.918)**	(10.727)**	(1.395)	(2.364)*
$(k_{t-1} - n_{t-1})$	0.053	0.023	0.054	0.129	0.075
	(0.474)	(0.270)	(0.764)	(1.014)	(0.827)
n_{t-1}	1.069	0.577	0.278	1.000	0.828
	(4.342)**	(3.040)**	(3.498)**	(4.217)**	(8.023)**
s_h	0.207	0.228	0.849	0.766	0.849
s_m	0.164	0.124		0.435	
s_y	1.950	0.774		1.802	2.314
r_{hw}	-0.166	0.048	0.100	-0.087	0.036
r_{my}	0.457	-0.027		0.591	
Quasi-concavity (proportion)	0.338	0.384			
Monotonicity (proportion)	0.674	0.721			
Log likelihood value	-324.57	-299.34	-1369.78	-1353.29	-1330.66
Number of firms	289	289	289	289	289

Note: All regressions include dummy variables for country, industry, ownership and time. The numbers in ( ) are *t*-statistics based on asymptotic standard errors. Significance at the one per cent, five per cent and ten per cent level is indicated by *, ** and [+] respectively.

[$] For the translog production function, the reported numbers associated with the inputs k_t, n_t, m_t, e_t are marginal effects. These are functions of the inputs, and have therefore been evaluated at sample means. The standard errors and *t*-values have also been evaluated at sample means.

TABLE 5

SELECTED OUTPUT PRODUCTION FUNCTION GMM ESTIMATES

	[1] Levels[1]	[2] First Differences[2]
y_{t-1}	0.086	0.041
	(2.787)**	(1.082)
$export_{t-1}$	0.060	0.163
	(1.735)$^{+}$	(1.116)
k_t	0.022	-0.023
	(1.657)$^{+}$	(1.193)
n_t	0.126	0.207
	(6.184)**	(4.212)**
m_t	0.675	0.620
	(27.320)**	(19.275)**
e_t	0.108	0.082
	(5.923)**	(2.768)**
Sargan-Hansen: (*J*, d.f., *p*)	(4.98, 5, 0.42)	(5.08, 4, 0.28)
Number of firms	289	289

Note: The numbers in ( ) are *t*-statistics. Significance at the one per cent, five per cent and ten per cent level is indicated by *, ** and $^{+}$ respectively. Hypothesis tests are based on robust, finite sample corrected standard errors calculated using the method proposed by Windmeijer (2000).

(1) The regression includes a constant and dummy variables for country, industry, ownership and time. The instrument set consists of a constant, Δy_{t-1}, Δk_t, Δn_t, Δm_t, Δe_t, $\Delta export_{t-1}$, k_t, n_t, m_t, e_t, $export_{t-1}$, dummy variables for country, industry, ownership and time.

(2) The regression includes a constant. The instrument set consists of a constant, y_{t-2}, k_{t-1}, n_{t-1}, m_{t-1}, e_{t-1}, $export_{t-1}$, k_t, n_t, m_t, e_t.

Appendix: Likelihood Functions Underlying the ML Models

Our ML estimators are similar to Keane et al. (1988) and Clerides et al. (1998). For notational simplicity, express the production function (1) as

$$y_{it} = z_{1it} b_1 + m_i + h_{it},$$

and the export equation (2) as

$$x_{it}^{*} = z_{2it} b_2 + y_i + w_{it},$$

where

$$x_{it} = \begin{cases} 1 & if\ x_{it}^{*} \geq 0 \\ 0 & if\ x_{it}^{*} < 0 \end{cases}$$

is observed. Conditional on m_i, y_i, z_{1it} and z_{2it}, the contribution of firm *i* to the sample likelihood is equal to

$$L_i = \prod_{t=1}^{T} \Phi\left((2x_{it} - 1) \cdot \left(z_{2it} b_2 + y_i + \left(r_{hw} / s_h \right) \cdot \left(y_{it} - z_{1it} - m_i \right) \right) \cdot \left(1 - r_{hw}^2 \right)^{-0.5} \right) \times s_h^{-1} \cdot f\left(\left(y_{it} - z_{1it} b_1 - m_i \right) / s_h \right), \tag{A1}$$

where $\Phi(\cdot)$ and $f(\cdot)$ are the standard normal distribution and density functions, respectively.

A. No unobserved heterogeneity

No unobserved heterogeneity of the form $s_m^2 = s_y^2 = r_{my} = 0$ implies that m and y are constant across firms. In this case the likelihood (A1) can be written

$$L_i = \prod_{t=1}^{T} \Phi\left((2x_{it} - 1) \cdot \left(z_{2it} b_2 + \left(r_{hw} / s_h \right) \cdot \left(y_{it} - z_{1it} b_1 \right) \right) \cdot \left(1 - r_{hw}^2 \right)^{-0.5} \right) \cdot s_h^{-1} \cdot f\left(\left(y_{it} - z_{1it} b_1 \right) / s_h \right). \tag{A2}$$

(see Clerides et al., 1996, Appendix III). The sample log likelihood, written as

$$\log L = \sum_i \log L_i(\cdot), \tag{A3}$$

is straightforward to maximise using some iterative method. For all ML results reported in the paper we use the SAS/IML NLPDD subroutine to maximise the log likelihood function.

B. The CLT model

Under the assumption that m and y follow a bivariate normal distribution, we can express the likelihood function conditional on observable data by integrating out m and y . To deal with the initial conditions problem arising from the presence of heterogeneity and dynamics, we follow Heckman's (1981a, 1981b) suggestion of adding equations to the system that model the initial conditions y_{i0} and x_{i0} as functions of exogenous regressors z_{3i0} and z_{4i0}, respectively, and the firm effects:

$$y_{i0} = z_{3i0}b_3 + t_3 \cdot m_i + x_{3i0}$$

$$x_{i0}^{*} = z_{4i0}b_4 + t_4 \cdot y_i + x_{4i0} \,,$$

where t_3 and t_4 are factor loading parameters, and the residuals are normally distributed:

$$(x_{3i0}, x_{4i0}) \sim N(0, \Lambda), \qquad \Lambda = \begin{vmatrix} s_{x3}^{2} & \\ 0 & 1 \end{vmatrix} .$$

The resulting individual likelihood can be written

$$L_i = \iint L_i\left(\cdot \middle| m, y\right) dF(m, y) \,, \tag{A4}$$

where

$$\begin{aligned} L_i\left(\cdot \middle| m, y\right) = \prod_{t=1}^{T} \Phi\Big(& (2x_{it} - 1)\cdot\left(z_{2it}b_2 + y + \left(r_{hw}/s_h\right)\cdot\left(y_{it} - z_{1it}b_1 - m\right)\right)\cdot\left(1 - r_{hw}^{2}\right)^{-0.5}\Big)\times \\ & s_h^{-1}\cdot f\left(\left(y_{it} - z_{1it}b_1 - m\right)/s_h\right)\cdot\Phi\left(\left(2x_{i0} - 1\right)\cdot\left(z_{4i0}b_4 + t_4 \cdot y\right)\right)\cdot s_{x3}^{-1}\times \\ & f\left(\left(y_{i0} - z_{3i0}b_3 - t_3 \cdot m\right)/s_{x3}\right), \end{aligned} \tag{A5}$$

and $F(.)$ is the bivariate normal distribution. To solve (A4) we follow CLT and use a bivariate Gauss-Hermite quadrature, which involves expressing m and y as linear combinations of two orthogonal random terms using a Cholesky decomposition, and then integrating over the two orthogonal random terms using standard (univariate) quadrature techniques (see e.g. Judd, 1998, Chapter 7). We then maximise the sample log likelihood using the SAS/IML NLPDD subroutine.

C. The NPML model

The discrete, multinomial equivalent of the CLT individual likelihood function (A4) is equal to

$$L_i = \sum_{q=1}^{Q}\sum_{r=1}^{R} P_{qr} L_i\left(\cdot \left| m = m_q, y = y_r\right.\right), \tag{A6}$$

where

$$\begin{aligned} L_i\left(\cdot \left| m = m_q, y = y_r\right.\right) = \prod_{t=1}^{T} & \Phi\left(\left(2x_{it} - 1\right)\cdot\left(z_{2it}b_2 + y_r + \left(r_{hw}/s_h\right)\cdot\left(y_{it} - z_{1it}b_1 - m_q\right)\right)\times\right. \\ & \left.\left(1 - r_{hw}^2\right)^{-0.5}\right)\cdot s_h^{-1}\cdot f\left(\left(y_{it} - z_{1it}b_1 - m_q\right)/s_h\right)\times \\ & \Phi\left(\left(2x_{i0} - 1\right)\cdot\left(z_{4i0}b_4 + t_4\cdot y_r\right)\right)\cdot s_{x3}^{-1}\times \\ & f\left(\left(y_{i0} - z_{3i0}b_3 - t_3\cdot m_q\right)/s_{x3}\right), \end{aligned} \tag{A7}$$

and

$$\sum_{q=1}^{Q}\sum_{r=1}^{R} P_{qr} = 1,\ P_{qr} \geq 0 \text{ for all } q,r.$$

The restrictions on the probability terms are imposed by specifying appropriate boundary and linear equality constraints in the computer code. Some trivial normalisations are also necessary (see Mroz, 1999). Because we include intercepts in each equation only Q-1 and R-1 support points are identified. Following Blau (1994) we parameterise the support points as $m_q = \Gamma_m \cdot W_{mq}$ and $y_r = \Gamma_y \cdot W_{yr}$ where Γ_m and Γ_y are scale factors and

$$W_{mq} = \begin{cases} -0.5 & \text{if } q = 1 \\ 0.5 - \left(1 + \exp\left(-a_{mq}\right)\right)^{-1} & \text{if } 1 < q < Q \\ 0.5 & \text{if } q = Q \end{cases},$$

$$W_{yr} = \begin{cases} -0.5 & \text{if } r = 1 \\ 0.5 - \left(1 + \exp\left(-a_{yr}\right)\right)^{-1} & \text{if } 1 < r < R \\ 0.5 & \text{if } r = R \end{cases}.$$

The sample log likelihood is maximised using the SAS/IML NLPDD subroutine. The estimation exercise is quite costly, since convergence is slow and may occur at a local optimum. To guard against convergence at local points we adopt a 'brute force' multiple step procedure, suggested by Thomas

Mroz in a personal communication. The first step is to take 50 bootstrap samples from the original sample, assign random start values, carry out 15 iterations from the random start values for each sample and store the resulting estimates. We proceed by using the original sample and carry out 15 iterations using each of the 50 estimates as start values. We then select the 25 parameter vectors associated with the highest log likelihood values, and try to bring each one to an optimum. The one with the highest function value is taken to be the maximum likelihood estimator. We experimented with increasing the number of bootstrap samples as well as the number of intermediate iterations, but found the above numbers to be adequate.